

UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF NEW YORK

-----X

UNITED STATES OF AMERICA

-against-

Docket No. 04 CR 1016 (S-2) (NGG)

RONELL WILSON,

Defendant.

-----X

**RONELL WILSON'S POST-HEARING REPLY MEMORANDUM
IN SUPPORT OF RELIEF PURSUANT TO
ATKINS V. VIRGINIA AND 18 U.S.C. § 3596(c)**

MICHAEL N. BURT
Law Office of Michael Burt
1000 Brannan Street Suite 400
San Francisco, California 94103

DAVID M. STERN
COLLEEN BRADY
Attorneys for Defendant
Ronell Wilson

BEVERLY VAN NESS, Of Counsel

TABLE OF CONTENTS

	<u>Page</u>
A. THE GOVERNMENT HAS IGNORED MYRIAD PROBLEMS WITH THE IQ SCORES IT RELIES UPON, INCLUDING THOSE ACKNOWLEDGED BY ITS OWN EXPERTS (replying to govt. brief, pp. 4-16 and <i>passim</i>)	1
B. THE BEST PRACTICES SET FORTH IN THE LATEST AAIDD MANUAL AND USER'S GUIDE MUST BE APPLIED, INCLUDING THE FLYNN EFFECT (replying to govt. brief, pp. 17-30)	7
C. THE GOVERNMENT MISCONSTRUES PRACTICE EFFECT INFLATION, AND WRONGLY ARGUES THAT THE LACK OF RAW DATA FOR MOST TESTS IS OF LITTLE SIGNIFICANCE SINCE THE SCORES ARE PURPORTEDLY CONSISTENT (replying to govt. brief, pp. 33-34, 36-38)	11
D. THE GOVERNMENT'S USE OF SNIPPETS OF CURRENT "EVIDENCE" TO SUPPORT A CLAIM THAT MR. WILSON IS ABLE TO THINK ABSTRACTLY IS MISPLACED (replying to govt. brief, pp. 45-48)	14
E. THE GOVERNMENT'S SUMMARY OF MR. WILSON'S SIGNIFICANT ADAPTIVE BEHAVIOR DEFICITS, WHICH ARE SPREAD OVER MANY DOMAINS, IS CURSORY, HIGHLY SELECTIVE, AND IGNORES THE CONCESSIONS OF ITS OWN EXPERTS (replying to govt. brief, pp. 39-44)	16
CONCLUSION	20

This memorandum is submitted in reply to the government's amended memorandum opposing Mr. Wilson's claim for *Atkins* relief (ECF #984, dated 1/7/13).

A. THE GOVERNMENT HAS IGNORED MYRIAD PROBLEMS WITH THE IQ SCORES IT RELIES UPON, INCLUDING THOSE ACKNOWLEDGED BY ITS OWN EXPERTS (replying to govt. brief, pp. 4-16 and *passim*).

The government has summarized the IQ tests administered to Mr. Wilson, arguing that virtually all – save for Dr. Nagler's 1994 test – should be credited (*see* ECF #984, pp. 4-16). Putting aside for the moment the inflation of scores due to the practice/progressive effects and the lack of raw data for most tests (making it impossible to cross-check for a variety of possible examiner errors), which will be discussed below, the government fails to acknowledge numerous facts that undercut the reliability of many of these IQ scores. For example, both Dr. Denney and Dr. Mapou conceded at the hearing that IQ tests results for youngsters – through age 10 for the former and pre-teen for Dr. Mapou – are not a reliable predictor of a person's intelligence in later life; a child's IQ score is not stable. Dr. James further explained that a child's frontal lobe network, which controls executive functioning, is insufficiently developed (*see* defense opening brief, ECF #982, at pp. 40-41 [detailing testimony]).¹ Mr. Wilson was not yet 7 when he had his first IQ test, and only 9 1/2 years old when he had his second test.

An additional problem with the second test – not mentioned in the government's brief – was Dr. Drezner's substitution of mazes for object assembly, the latter being a core subtest for the performance IQ. This was a departure from procedure according to Dr. James, who also testified that mazes tested something completely different from object assembly (*see* ECF #982,

¹ Page numbers refer to those set forth at the bottom of each page of the memorandum, which correspond to the Table of Contents, not to the ECF-filed page numbers at the top.

p. 42). These assertions that were not disputed by Dr. Denney.

The government also whitewashes Mr. Giglio's testimony. For example, it inflates Mr. Wilson's academic improvement while at Brookwood – he actually improved very little, and was “below average” even among other youths confined for criminal offenses despite “very close intensity teaching”; the government also ignores Giglio's continued recognition of Mr. Wilson's intellectual deficiencies, and the latter's inability to think abstractly even after two years in a structured and therapeutic environment (*compare* defense opening brief, pp. 43-44, with govt. brief, pp. 9-10, including argument that Mr. Wilson's supposed educational advancement at Brookwood yielded a higher score during testing by Mr. Popp in 2000).

Mr Popp's undisputed prorating of a core subtest is dismissed as insignificant, as if the testing manuals approve of such a procedure (*see* govt. brief, pp. 34-35). In fact, the WAIS-III manual provided instructions for pro-rating only as a last resort: if a core subtest was spoiled or impossible to administer, an alternative subtest was supposed to be used. Here, Popp did not explain why he did not administer Picture Arrangement, a performance IQ test, and he did not give an alternative subtest of Symbol Search or Object Assembly. Instead, he prorated, which Dr. Shapiro and Dr. James testified likely inflated Mr. Wilson's score (*see* defense opening brief, pp. 25-27). It certainly made Popp's score less reliable, as Dr. Mapou, a government expert, explicitly conceded: a prorated score is merely *an estimate*, and, therefore, necessarily less accurate (H. 2044).²

² The government argues, using a chain of inferences, that Dr. Nagler also prorated a subtest (govt. brief, p. 7). The fact remains, however, that there is no acknowledgment of pro-rating anywhere in Nagler's report or raw data as there is in Popp's report (*see* evidence in his report cited in defense brief, p. 26; there is no raw data for Popp's test). Moreover, none of the government experts ever suggested Nagler had prorated. The prosecutors did not question them

As for the three scores for which raw data is available, the government is similarly selective and, with respect to Dr. Drob's 2003 test, misleading. We have already discussed, at length, why Dr. James properly relied on Drob's testing as generating a reliable score which this Court should credit: the test's close proximity to the crime, when Mr. Wilson was a young adult; Drob's administration of a full battery of neuropsychological tests, the results of which were consistent with Dr. James's tests (no other tester gave this battery, in whole or in part); the availability of the raw data to cross-check; and the doctor's opinion that Mr. Wilson gave his best effort throughout the testing process (*see* defense opening brief, pp. 45-46). Drob also frankly acknowledged procedures he did *not* adopt in 2003 which would have made a difference. For example, he was not familiar with the Flynn effect (which Dr. Denney acknowledged was a real phenomenon) and made no adjustment for it, he did not compute a confidence interval (as Dr. James *and* Dr. Denney did), and he did not consider practice/progressive effects (again, acknowledged as a legitimate inflationary factor by Dr. Denney for as long as 12 to 13 years) (*see* defense brief, pp. 37-40, 45-46).

The government did not acknowledge any of this (*see* govt. brief, pp. 12-14). Instead, it harps on Drob's prior conclusion – without the benefit of the adjustments cited above – that Mr. Wilson was not mentally retarded, and unfairly insinuates that Drob's IQ results were artificially low because Mr. Wilson was malingering (*id.*, p. 14). This is a unquestionable distortion of Dr. Drob's testimony. He described the particular tests he administered which were designed to expose malingering, and was emphatic, to a reasonable degree of scientific certainty, that Mr. Wilson put forth his best effort in all aspects of testing, both on the WAIS-III nad on RBANS

about the point.

tests he had also conducted (H. 916-21).

On cross, the prosecutor referred Drob to a test given to Mr. Wilson in 2005 (two years after the WAIS-III Drob had administered), in conjunction with a PET scan. Because this test was similar to one of the several RBANS tests that Drob had described in his “best effort” testimony on direct, and because Mr. Wilson had done better in 2005 on this one test than in 2003, the prosecutor tried to suggest that Mr. Wilson had failed to put forth his best effort on the IQ test. But Dr. Drob never wavered from his original opinion (*see* H. 1021-29).

Citing only portions of the cross-examination, the government now argues that perhaps Mr. Wilson was malingering on the 2003 IQ test, and that his score was artificially low as a result (govt. brief, p. 14). This suggestion is without any evidentiary support, and is in fact belied by Dr. Drob’s unequivocal testimony to the contrary. It also contradicts the government’s earlier statement, when seeking to undercut Dr. Nagler’s test results, that “the best person to determine what weight to give a score is the psychologist who administered the test and who was able to observe the test-taker’s actions” (govt. brief, p. 6). Dr. Drob had notes, his report, and all of his raw data to review, and he also recalled testing Mr. Wilson. He was certain that Mr. Wilson gave his best effort in 2003, and nothing undercuts that opinion.

The government also seeks to discount Dr. Nagler’s test results because of a lack of best effort on Mr. Wilson’s part (govt. brief, p. 6). Although in this case, there is some evidence in Nagler’s notes to support this – as we acknowledged in our opening brief (p. 46) -- Dr. Nagler did not terminate the testing or indicate that the results were invalid, and her raw data demonstrates that Mr. Wilson completed all of the subtests that were presented (*see* Dr. James, H.1450-51). On the contrary, Dr. Nagler observed that, despite his squirming and yawning and

generally careless approach, “[n]evertheless, there were occasions when he appeared interested and concerned with doing well. However, as soon as the tasks became somewhat difficult, he became frustrated and gave up” (Defense Exhibit C5, Nagler IQ test, at GOV003942).

It is also important to acknowledge that Nagler’s was the third administration of the WISC-III, all within a three-year period and only a little over 13 months from the second WISC-III Mr. Wilson had taken. Inflation of scores due to practice effects cannot be ignored (*see* defense opening brief, pp. 30-33, 38-40, and discussion below). Dr. James explained that although practice effects inflate a score, the effects are not always visible and can be offset by other factors such as examiner error (defense brief, p. 39). Thus, as Dr. James testified, even if Mr. Wilson did not provide his best effort, 100%, throughout all aspects of Dr. Nagler’s testing, any reduction on his overall score should be compensated for by an increase due to practice effect (H. 1451-52). In other words, his FSIQ score of 70, Flynn-adjusted to 68.35, should be deemed reliable, with practice effect inflation offsetting less than optimal effort.

Practice effects also inflated Dr. Denney’s test results in 2012 – this was the ninth administration of a Weschler instrument to Mr. Wilson. It was also nine years after the crime, all that time being spent in a structured environment, which Dr. Denney admitted could operate to increase Mr. Wilson’s IQ. Further, the WAIS-IV has eliminated speed tests, which Dr. Shapiro testified could also be expected to increase Mr. Wilson’s score (defense brief, pp. 46-47).

We also note that Mr. Wilson’s Flynn-adjusted score in 2012 is 78.02, only three points above a qualifying score even if an absolute cut-off number was required – which it is not – and

the lower band of the 95% confidence interval is 73.70.³ Under such circumstances, seizing on this IQ number to rule out prong one would not only be contrary to the full record, but would result in an arbitrary and unreliable determination of the *Atkins* exemption from execution.

Dr. James's opinion that Mr. Wilson meets prong one is based on a host of factors, including his myriad adaptive behavior deficits (discussed below), which the government agrees can inform the intellectual functioning prong (govt. brief, p. 30). Mr. Wilson's Flynn-adjusted scores were all very low, Dr. James testified, low enough to require her consider case-specific

³ The government also urges that an individual's true IQ, when dealing with low obtained scores -- scores at or near the range that would satisfy the first prong of the definition of intellectually disabled -- most likely falls toward the upper end of the confidence interval, as a result of the statistical phenomenon known as regression to the mean (govt. brief, p. 32). And the government, unfairly, suggests that Dr. James agreed that, because of regression to the mean, Mr. Wilson's "true IQ" likely falls toward the upper end of the of the confidence interval for each of the obtained scores. Dr. James, however, explicitly did not agree (H. 1457: "I don't know that that's correct."). Neither the AAIDD nor the APA endorse attempts to predict where, within the reported confidence interval, a so-called "true IQ" falls: both, in their definition of the disability, simply require that, given the imperfect nature of instruments used to measure IQ, obtained scores be treated as a range. *See AAIDD 2010 Manual*, at 36 ("Both AAIDD and the American Psychiatric Association (2000) [author and year of publication of the DSM-IV-TR] (2000) support the best practice of reporting an IQ score with an associated confidence interval."); *DSM-IV-TR*, at 41-42. Indeed, the AAIDD has expressly rejected Dr. Denney's approach to calculating "true IQ," precisely because of the problem of regression to the mean, which the AAIDD identified as a problem, particularly for those at the low end of the IQ spectrum, and not as a virtue:

It is tempting to suggest that examiners use a true score correction when classifying mental retardation; however, there is a significant limitation to Stanley's formula. [Stanley, J. (1971). Reliability. In R. L. Thorndike (Ed.), Educational measurement (2nd ed., pp. 50-72). Washington, DC: American Council on Education.] It will always yield a score that regresses to the mean. Because of this, it will always yield higher estimates of intellectual functioning for people who have mental retardation, and this is not suitable for the diagnosis of mental retardation.

AAMR 2002 (10th ed.) Manual, at 59. Dr. Denney's apparent lack of familiarity with the AAIDD's rejection of his approach underscores his lack of experience in the field of intellectual disability.

issues like Mr. Wilson's performance on her neurological tests, practice/progressive effects based upon multiple administrations of a Wechsler instrument over the years, the existence or lack of raw data for the IQ tests, prorating, and the confidence interval bands around his scores (*see, e.g.*, H. 1470-71).⁴ Dr. James used her clinical judgment -- grounded firmly in all of the available data, the best practices endorsed by the leading professional organization, and her own years of working, evaluating, and diagnosing individuals with ID/MR -- to make her diagnosis.

B. THE BEST PRACTICES SET FORTH IN THE LATEST AAIDD MANUAL AND USER'S GUIDE MUST BE APPLIED, INCLUDING THE FLYNN EFFECT
(replying to govt. brief, pp. 17-30)

The government asks the Court to ignore the AAIDD's current standards for assessing intellectual disability (brief, pp. 17-26). It would divorce the bare definition of MR/ID from the explicative material in the AAIDD Manual and User's Guide, which informs the definition and reflects the best practices for determining the issue – best practices as promulgated by the leading professional association in the country. This is akin to asking the Court to blind itself to advances in DNA analysis, however well-established and scientifically-accepted current research practices are. The government does not cite one case to support its position, and does not

⁴ In addition to a full battery of neurological tests, Dr. James administered a set of achievement tests to Mr. Wilson. The government asserts that the latter results corroborate Dr. Denney's opinion that Mr. Wilson is not mentally retarded, relying on some isolated comments by Dr. Mapou – who has no qualifications as a mental retardation expert (*see* H. 2027-30, 2070-73) (govt. brief, pp. 38-39). Dr. James herself explained, at length, that achievement testing addressed academic functioning – reading, spelling and the like – and that she gave these tests to determine if Mr. Wilson suffered only from a learning disability. Not only did the 30-year old defendant score quite low – from a second grade level to seventh, with most scores in the third, fourth and fifth grade level – but his past and current academic tests revealed a broader range of conceptual deficits than would be present if he had only a learning disability: this was demonstrated consistently over time, starting in elementary school. The prosecutor never asked its IQ expert, Dr Denney, about the achievement tests (H. 1422-32; *see also* H. 1409-14).

acknowledge the many cases we have cited against it (*see* defense brief, pp. 9-10).

The government's request, in short, lacks both logic and supporting authority. It presses its claim by mislabeling the AAIDD's current standards as "codify[ing] some of the most controversial topics" at issue in this case: "the 2010 Manual takes strong positions against considering past criminal, custodial or verbal behavior in assessing adaptive functioning [*see* discussion in Part D. below]; discourages consideration of cultural or socioeconomic factors as they may affect IQ-testing performance; and expressly endorses the application of the Flynn effect" (govt. brief, pp. 17-18). But these guidelines are not "controversial," they are based on documented research studies and evolving professional standards and practices.

The Flynn effect, for example, is a well-established phenomenon based on decades of research. *See, e.g., United States v. Hardy*, 762 F.Supp.2d 849, 857-63 (E.D. La. 2010) (discussing research and case support); *United States v. Davis*, 611 F. Supp.2d 472, 485-87 (D. Md. 2009) (same). As Dr. Flynn himself observed, "Failure to adjust IQ scores in the light of IQ gains over time turns eligibility for execution into a lottery – a matter of luck about what test a school psychologist happened to administer." *Davis*, at 486. A failure to make a Flynn adjustment has the following result: "one person will meet the criterion of mental retardation, and another person will be judged not to have done so, purely because one took a test with current norms and the other took a test with obsolete norms." *Id.*

The government cites a handful of cases which allegedly demonstrate "the controversial nature of the Flynn effect" (govt. brief, pp. 26-29). They are all habeas cases, which turn on whether a state court's failure to apply the effect was contrary to or an unreasonable application of clearly-established federal law. *See* defense brief, p. 29, n.5, acknowledging such cases. The

defense has cited many more cases that have applied the Flynn effect, including those in which federal District Court judges were considering *Atkins* claims as the initial fact-finder, and several Circuit Court opinions endorsing Flynn (defense brief, pp. 27-29). These decisions are consistent with the scientific literature and opinions of experts in the field, as well as with the expert testimony in this case – including that of Dr. Denney, who acknowledged the validity of the Flynn effect (*see* testimony cited in defense brief, pp. 30, 38).

The government also repeatedly asserts that Mr. Wilson's obtained scores underestimate his potential. It urges this Court to follow the path of its experts and, in effect, inflate Mr. Wilson's obtained scores, to award bonus points, because of emotional and socioeconomic or “cultural” factors (*see* govt. brief, pp. 3, 8, 9, 31, 42) -- a word that the government largely leaves undefined, but which suggests, as was explained at the hearing, not some unusual or little-understood background (from the perspective of American examiners), but rather behaviors or experiences that the government and its witnesses associate with Mr. Wilson's upbringing in an impoverished African-American family and community in New York City in the 1980s and 1990s (Drezner, H. 1734).

As the government concedes, however, the AAIDD Manual “discourages consideration of cultural or socioeconomic factors as they may effect IQ-testing” (*id.*, p. 18). As Dr. James explained, IQ tests are normed “on individuals across the country from different regions from different ethnic backgrounds and different socioeconomic backgrounds” (H. 1435). Race, culture, and socioeconomic considerations are addressed through the sample groups on which the tests are normed, and neither the AAIDD nor any other protocol permits “discount[ing]” or “dismiss[ing]” an obtained IQ score because of socioeconomic or cultural or racial

considerations (H. 1435-36; *see also* Dr. Olley, H. 491-92, 494).⁵ Indeed, the AAIDD identifies numerous “socioeconomic” (and perhaps “cultural,” as used by the government) issues as risk factors for intellectually disabilities. See AAIDD 2012 Manual, at 60, Table 6.1 (poverty, impaired parenting, inadequate family support).⁶ The corrections/inflations/bonus points that the government suggests should increase Mr. Wilson’s obtained scores would, in effect, factor out issues in his life that may have contributed to his condition. Worse, under the government’s approach, an African-American from an impoverished background would have to obtain an IQ score well below 70 in order to obtain the benefits of Atkins. That cannot be the law; and it certainly is not consistent with the best practices established by the AAIDD and APA.

Finally, the government attempts to justify its experts’ subjective upward correction by reference to “clinical judgment.” But neither the government nor its experts offer anything -- no corroborative testing or empirical data, no peer-reviewed academic scholarship, no confirmation or endorsement from either of the two professional organizations that set the standards for best

⁵ So, too, are the adaptive functioning instruments normed on the general population, including relevant racial, cultural, and socioeconomic groups. Impairments in adaptive functioning in Mr. Wilson’s community, then, serve as an additional check that his obtained IQ scores do, in fact, demonstrate real impairments in intellectual functioning, and not biases in the testing.

⁶ For the same reasons, Mr. Wilson’s obtained scores should not be discounted or inflated because of his “emotional problems.” Again, the AAIDD identifies, as a risk factor that might contribute to the disability, “difficult child behaviors.” AAIDD 2010 Manual, at 60 Table 6.1. Moreover, as both Drs. James and Olley explained, one may have both an intellectual disability and another condition (such an antisocial personality disorder) simultaneously (“dual diagnoses”): a second condition does not rule out a diagnosis of ID/MR (Olley, H. 498-92; Shapiro, H. 341-42). The government’s approach, rejected in the field, is to factor out potential contributing factors. The government also may confuse the cart and the horse, explaining Mr. Wilson’s ID/MR by reference to his behavior, instead of the other way around (*see* Shapiro, H. 341: “You may see behavioral disturbance arising primarily from the intellectual disability, yes.”).

practices in the field -- to which to tether or evaluate the exercise of this “judgment.” Instead, they reject the data based on nothing more than a subjective belief that the obtained scores are lower than Mr. Wilson’s potential. But this “I know it when I see it,” nothing more than “‘gut’ estimates” as characterized by Dr. James, H. 1365, is not the “clinical judgment” that is the focus of an entire chapter of the AAIDD manual. See AAIDD 2010 Manual, at 85-103. “Clinical judgment” must be based on “data, not just on your gut” (Dr. James, at H.1369).

C. THE GOVERNMENT MISCONSTRUES PRACTICE EFFECT INFLATION, AND WRONGLY ARGUES THAT THE LACK OF RAW DATA FOR MOST TESTS IS OF LITTLE SIGNIFICANCE SINCE THE SCORES ARE PURPORTEDLY CONSISTENT (replying to govt. brief, pp. 33-34, 36-38)

The government suggests that unless a person’s performance IQ score goes steadily up, practice/progressive effects may not be considered. In other words, the government selectively ignores the several tests for which Mr. Wilson’s performance IQ *did* rise – in 1993, 1998 and 2012 – and cites the years in which it went down as “evidence” that practice effects *never* inflated Mr. Wilson’s scores (govt. brief, pp. 36-37). But Mr. Wilson does not have to show a “consistent, steady increase in scores (*id.*, p. 36) in order for the Court to make a practice effects adjustment. Dr. James testified that these effects were variable, and are not always visible (*see* defense opening brief, p. 39).

Test-retest inflation was specifically acknowledged by Dr. Denney, even after a very long hiatus between administrations, and Denney also agreed that progressive error, based on multiple administrations of a testing instrument, was “a reasonable hypothesis” (*id*; *see* H. 1981, 1983). Dr. Shapiro and Dr. Mapou also endorsed progressive error (defense opening brief, p. 40). So, too, does Dr. Kaufman, “especially . . . if the person has been administered a Wechsler scale (*any* Wechsler scale) several times in the course of a few years” (*id.*, p. 32, citing Defense Exh. B).

Denney testified that Dr. Kaufman was a leading authority on Wechsler testing instruments (H. 1925-26). Denney was “not sure” whether there was “solid data” to support the progressive error phenomenon, but Kaufman cites studies in his discussion of the concept (see H. 1983-84, and Defense Exh. B).

Although not acknowledged by the government in its brief, Dr. Denney agreed with Dr. James that practice/progressive effects can be offset by other factors (H. 1929-31). The fact that a score did not rise, therefore, does not mean that practice effects inflation did not occur. And Denney’s “proof is in the pudding” quotation in the government’s brief (pp. 36-37), which it uses as the sole supporting evidence that practice effects can be considered *only* if scores steadily rise, is misleading at best. As Denney admitted at the hearing, that testimony was directed to practice effects exhibited in a single test/retest situation, *not* to the phenomenon of progressive error over the course of multiple administrations of a Wechsler instrument (H. 1927-29).

There was, in short, no testimony that practice/progressive effects can only be demonstrated by a consistent rise in scores, or that an adjustment cannot be made for these effects if scores go up and down over time. Instead, there was undisputed evidence from Dr. James, as endorsed by Dr. Kaufman’s treatise, that practice/progressive effects inflate scores even after long intervals, when a Wechsler test is administered repeatedly. Here, Mr. Wilson was given nine Wechsler IQ tests: a WISC-R, five WISC-III’s, two WAIS-III’s and a WAIS-IV. This was, as Dr. James stated without dispute, an “extraordinary” number (H. 1448).

Moreover, as the testimony made plain, the AAIDD eschews a hard cut-off IQ score for the intellectual functioning prong precisely to permit the evaluators to take into account data- or

research-driven concerns such as practice/progressive effects. If there were a cut-off, though, Dr. Denney's own Flynn-adjusted score of 78.02 lies a mere three points above the line. Best practices demand taking practice/progressive effects into account when weighing Dr. Denney's score, obtained years after the crime at issue here, during which Mr. Wilson was held in a structured setting, which Dr. Denney himself acknowledged could also inflate his scores. Because those considerations all work to push the scores higher, Denney's score cannot rule out a diagnosis of MR/ID. Drob's qualifying score, close in time to the crimes, on a test where Mr. Wilson gave best efforts and where the raw data permits the conclusion that no scoring errors were made, also suffers the inflating pressure of practice/progressive effects. Since the score satisfies the intellectual functioning prong, however, those effects are not significant. Drob's score should be taken as the ceiling, then, especially given all of the other corroborating evidence in the case -- in particular, the evidence of real-world limitations in adaptive functioning which serves to inform Mr. Wilson's obtained IQ scores.

The government also seeks to minimize the importance of missing raw data for many tests, precluding checks for a host of examiner errors, by arguing that the scores' consistency means they are correct.(govt. brief, pp. 33-34). However, the non-Flynn adjusted scores which the government relies on are *not* "consistent." There is a 14-point spread in Mr. Wilson's full-scale IQ, eight points without Nagler's score which the government would discard (*see* Dr. James' chart, attached as an exhibit to the defense opening brief). Since a person's IQ is supposed to be relatively stable over time, this significant differential must be explained by other factors, such as examiner error, practice effects, or the failure to adjust for obsolete norms. The components of Mr. Wilson's scores also vary widely (*see* chart): his verbal IQ ranges from 65 to

81, and his performance IQ from 80 to 95. The pairings are also quite variable, *e.g.*, 79/81, 65/80, 70/95, 80/92.

This variability also demonstrates that an IQ score is merely an estimate of a person's true intelligence as Dr. James testified (H. 1223, 1441). This is why a confidence interval must be applied (H. 1223-24). It is also why mental retardation expert – as Dr. James is and as the school psychologists who administered the historical IQ tests were not – must consider an array of other information to inform her clinical judgment and diagnosis.

D. THE GOVERNMENT'S USE OF SNIPPETS OF CURRENT "EVIDENCE" TO SUPPORT A CLAIM THAT MR. WILSON IS ABLE TO THINK ABSTRACTLY IS MISPLACED (replying to govt. brief, pp. 45-48)

The government devotes several pages of its brief to a handful of recent emails and phone calls sent by or involving Mr. Wilson. It argues that these illustrate his current ability to think abstractly, a sign of intelligence. But Dr. Shapiro made clear that even young children have some ability to abstract, and no one suggested at the hearing that a person with mild mental retardation has no capacity at all to think abstractly (H. 286). However, the AAIDD manual specifically instructs against using verbal behavior (as well as past criminal behavior) to make a mental retardation assessment (*see, e.g.*, H. 80-81). People at mild end of the mental retardation spectrum have reasonable language skills (Dr. Shapiro, H. 71, 78-80, 360-62). And on a superficial level, the person's difficulties in reasoning will not be apparent (H. 390-92).

More specifically, the manual precludes the use of verbal (and criminal) behavior for diagnostic purposes because there is not enough information to make an assessment and because there are no norms against which to compare the person's speech (H. 359-62, 393-94). As Dr. Olley explained, IQ tests and adaptive behavior scales have norms – "you can compare the

[individual's] performance to the performance of others of the same age.” But there are no norms for verbal discourse. In addition, the isolated bits of verbal expression in Mr Wilson’s emails and phone calls lack sufficient context to be of any clinical value (H. 831-32, 845-46). Thus, an opinion of Mr. Wilson’s intellectual capacities based on such “evidence” would be purely subjective (H. 845-46). And it could well be influenced by each assessor’s stereotype of how a mentally retarded person would or would not talk or act. What is required, instead, is an objective assessment based on normed testing instruments and well-established clinical standards.

Throughout its discussion of the IQ testing, the government also cites Mr. Wilson’s scores on the similarities subscale of the Wechsler (WISC and WAIS) IQ tests which, it asserts, “specifically tests abstract reasoning” and is thus, according to government, very important in the diagnosis of MR/ID (govt. brief, pp. 3, 4, 9, 11, 13). But the intellectual functioning prong of both of the relevant clinic standards is defined by reference to a full scale IQ score, not the score on an individual subscale. AAIDD 2010 Manual, at 31; DSM-IV-TR, at 41-42; *see also* Dr. James at H. 1237-38 (“the full scale IQ is the best estimate overall”); 1274 (“You can’t make a scaled score equivalent to an IQ score”); *id.* (“You don’t use a scaled score in order to determine an IQ, not in and of itself. So you can’t look at a scaled score and say, well, that’s a four. That’s an IQ of 70.”).

In any event, as with all of the scores on the Wechsler tests, the results on the similarities subscale are inflated by practice/progressive effects. For example, a close look at the individual questions of the similarities subscale on Dr. Drob’s administration of the WAIS-III and Dr. Denney’s administration of WAIS-IV show significant overlap in the questions (several

questions are repeated essentially verbatim or with essentially the same structure). *Compare* Defense Exh. M, Drob WAIS-III, at GOV004035, *with* Government Exh. 81, Denney WAIS-IV (raw data), at page 13 (titled “Similarities”). For each question, the respondent is awarded 0, 1, or 2 points, with 2 representing full credit. On the questions that are basically the same, Wilson was awarded full points; on questions that were completely new (with the exception of question 11), Wilson was awarded zero or one point. In other words, he did well on the familiar questions, but worse overall (on the new questions) -- exactly the kind of inflation described by practice/progressive effects. And each of Wechsler instruments at issue in this case, from the WISC-R to the WAIS-IV, included a similarities subscale, so that the practice/progressive effects began in 1989 with the first administration.

E. THE GOVERNMENT’S SUMMARY OF MR. WILSON’S SIGNIFICANT ADAPTIVE BEHAVIOR DEFICITS, WHICH ARE SPREAD OVER MANY DOMAINS, IS CURSORY, HIGHLY SELECTIVE, AND IGNORES THE CONCESSIONS OF ITS OWN EXPERTS (replying to govt. brief, pp. 39-44)

Although not acknowledged in the government’s brief, all three of its experts admitted that Mr. Wilson had significant adaptive behavior deficits in at least two domains listed by the DSM. This included Dr. Patterson (*see* H. 1846-51),⁷ whom the government cites as an authority (brief, pp. 40-41), even though he had absolutely no qualifications as an expert in mental retardation and was not offered as such (*see* H. 1739-65). The government does not mention Dr. Mapou, who acknowledged deficits in three domains – academic, social and communication – and believed Mr. Wilson probably had a deficit in the health and safety domain as well (H. 2064-68). Dr. Denney, the only government expert qualified to opine on mental retardation, did not

⁷ Patterson was only asked about two domains, academic skills and social and interpersonal skills, and acknowledged Mr. Wilson had demonstrated significant deficits in both.

concede simply that Wilson had “some deficits” in the academic and social domains, as the government states (brief, p. 40). Rather, he conceded significant limitations in both those areas, as well as in communication (H. 1938-39, 2018-25). He also acknowledged deficits, but “less so,” in home living and self-care (H. 1939).

The issue for Dr. Denney was not *whether* Mr. Wilson had significant deficits in two or more domains – in which case he would meet prong two – but why. Denney thought causation was a necessary component of a mental retardation diagnosis, even as he agreed that his opinion was directly at odds with the AAIDD (H. 1938-39, 2020). He did not offer any support for his position, no scientific literature or journal articles, no mental retardation experts who similarly bucked the judgment of the leading manual in the filed, no analysis -- it was simply his opinion.

Moreover, Dr. Denney’s opinion was contradicted by the evidence. First, he expressed the view that the deficit in academic skills was the result of a learning disability, which also bled into the deficit in communications (*seei* H. 1938, 2019). He acknowledged that Mr. Wilson was still “significantly behind” academically when he left Brookwood as a teenager – despite the intensive help he had received there in a structured environment, where he was “forced to take classes” as the government put it in their brief (p. 40). At that point, Dr. Denney testified, Mr. Wilson had started to outgrow his learning disability, he only had “residual signs of it” (H. 2011-22). In fact, Dr. Denney testified, Wilson “seems to have outgrown” the learning disability by age 18 (H. 1894, 2017-19, 2023). A lot of his academic difficulties in his later teens -- but not earlier, Denney conceded -- were supposedly the result of Wilson’s “unwillingness to apply himself in school” rather than intellectual limitations (H. 2021-22).

However, by the time Denney was called into this case, Mr. Wilson had been

incarcerated for nine years in a structured environment. This included many years on death row, where he had sought to improve himself and tried to obtain his GED. Significantly, when Dr. James gave him achievement tests in 2012 to measure his reading, spelling and math skills, he again demonstrated severe deficits on a par with his earlier history:

Mr. Wilson continues to demonstrate significant academic weaknesses on both basic (i.e. spelling) and applied (i.e. essay composition) tasks. In the current evaluation, single-word reading for nonsense and real words was measured in the mid 3rd grade to mid 4th grade range, with reading comprehension at the late 2nd grade level. Written expression, including spelling, sentence composition and essay composition, was measured in the early 4th grade to early 6th grade range. His essay was one paragraph and contained few elaborations of his ideas. Mathematics skills, encompassing calculation, math problem solving (i.e. applied mathematical knowledge) and fluency, were in the early 4th grade to early 5th grade range, with a relative strength on simple, single-digit addition (early 7th grade level). Notably, Mr. Wilson seemed to struggle with a particular aspect of single-digit subtraction, consistently producing an incorrect response on items involving subtracting zero.

(James report, p. 17). Clearly, Mr. Wilson's significant deficits in academics and communication have not diminished, and cannot be dismissed as resulting from a learning disability or a lack of effort – even if causation was relevant which it is not.

Dr. Denney also sought to explain Mr. Wilson's deficits in social skills as resulting from "wilful conduct." Yet he testified that this had nothing to do with Mr. Wilson's deficits in this domain before his late teens -- he could not pinpoint any dividing line -- and waffled repeatedly on his point. On examination by defense counsel, he said he was not necessarily opining as counsel tried to summarize his view, but he never explained what he *did* mean (*see* H. 2019, 2023-25). Moreover, the focus of prong two is not on a person's criminal behavior (*see* govt. brief, 41-42), which the AAIDD states should not be considered. Equally important, prong two involves a review of 11 discrete categories of behavioral skills, none of which, on their face,

concerns criminal behavior. Indeed, such activities are largely if not totally irrelevant to these domains, all of which test skill sets as demonstrated in a wide array of contexts. A mental retardation assessment does not address criminal behavior, or “maladaptive behavior.” It addresses adaptive behavior across many areas of living in the community. And, as the defense experts testified (Shapiro, H. 61-63; Olley, H. 471, 479-80), it concerns significant *deficits* in behavior, which cannot be cancelled out by isolated strengths.

Finally, the government has ignored the voluminous records concerning Mr. Wilson’s life, including a chart summarizing relevant portions around the adaptive behavior domains (Exh. A, cited in Mr. Wilson’s opening brief, p. 49). These records reveal a boy and teen who cannot express himself, who struggles with basic tasks, who acts inappropriately with his peers despite his obvious need to be accepted, and who presents as a person much younger than his actual age. The government also ignores the wealth of anecdotal reports from people who knew Mr. Wilson in different periods of his developmental period, as reflected in Dr. Olley’s report and testimony cited in our opening brief. The government sweepingly dismisses the interviews as “statements of biased friends and family” (brief, p. 42). These are the people who knew Mr. Wilson, of course, and they thus had the information needed.

Moreover, there is no evidence that any of them were biased, in the sense of deliberately overstating Mr. Wilson’s deficits in order to help secure a mental retardation diagnosis. Dr. Olley interviewed a dozen people, not simply the four to whom he administered the ABAS test, and their credibility was not suspect. Dr. Denney did not suggest anyone was biased, either. At most, he said that Mr. Wilson’s sister, Depetra, although not misrepresenting his behavior, may have been influenced by her view that he needed taking care of (Denney report, p. 47). But she

has known her brother all his life, and there is no question that he needed help throughout and that it was sorely lacking in his home environment. Denney also stated that Monica Cook *perhaps* did not view Mr. Wilson objectively, since emails suggested she had conflicting feelings about him (*id.*). However, this would not affect her factual reports about his past behavior, *e.g.*, his inability to prepare simple foods and even to remember to turn off the stove, and his generally poor judgment about basic household matters (Olley report, pp. 15, 17). Notably, Mr. Wilson was a young adult during the two years he lived with her, directly preceding the capital crimes. Her reports were a key component of Dr. Olley's assessment of Mr. Wilson's significant deficits in several domains, including home living.

CONCLUSION

For all the reasons stated, herein and in defendant's opening brief, Mr. Wilson respectfully asks this Court to conclude that he is intellectually disabled and thus ineligible for the death penalty.⁸

⁸ Dr. James testified about the an amendment to the Individuals with Disabilities Act which disfavors the use of discrepancy analysis in diagnosing learning disabilities (H. 1459-60). The Court requested a citation: the Act is at 20 U.S.C. Section 1400 *et seq.* The amendment was in 2004, and here is a brief description: "IDEA 2004 attempts to replace the discrepancy analysis with a standard that arguably is a more inclusive way to determine SLD in some children. For example, the new standard could qualify students who previously did not meet the required discrepancy. An example would a very low achieving student who also tested in the low range intellectually and thus did not qualify under a discrepancy analysis because there was no discrepancy between ability and achievement. On the other hand, the new standard also could exclude students who are very "bright" and only achieving academically at average or above average levels but not at their academic potential. Although too soon to assess how these changes will actually affect a student's eligibility, in some ways the new standard provides professionals greater flexibility but without the "bright-line" decision-making guidelines under discrepancy analysis." Bulletin #10 of the Special Education Law Quarterly, July 2007, at <http://wea.uwctds.washington.edu/HTML%20Bulletins/Bulletin10.html>.

CERTIFICATE OF SERVICE

I hereby certify that a copy of the foregoing Post-Hearing Reply Brief was served electronically on January 16, 2013, by filing a copy of same through the court's electronic filing system.

/ s /

MICHAEL N. BURT